УДК 004.5: 621

В.Н. Кулаковский, В.В. Цегельнюк, И.В. Скворцов
Институт сверхтвёрдых материалов им. В.Н. Бакуля, Украина
Украина, 04074, г. Киев, ул. Автозаводская, 2

ИНФОРМАЦИОННАЯ СИСТЕМА УПРАВЛЕНИЯ НАУЧНЫМИ ДАННЫМИ В ПРЕДМЕТНОЙ ОБЛАСТИ «СВЕРХТВЕРДЫЕ МАТЕРИАЛЫ»

V.N. Kulakyvskiy, V.V. Tsegelniuk, I.V.Skvortsov
V.N. Bakul`s institute for Superhard materials, Ukraine
Ukraine, 04074,c. Kyiv, Avtozavodskaya str, 2

MANAGEMENT INFORMATION SYSTEM OF SCIENTIFIC DATA IN THE SUBJECT AREA "SUPERHARD MATERIALS"

В.М. Кулаківський, В.В. Цегельнюк, І.В. Скворцов
Інститут надтвердих матеріалів ім. В.М. Бакуля, Україна
Україна, 04074, м. Київ, вул.. Автозаводська, 2

ІНФОРМАЦІЙНА СИСТЕМА КЕРУВАННЯ НАУКОВИМИ ДАНИМИ У ПРЕДМЕТНІЙ ГАЛУЗІ «НАДТВЕРДІ МАТЕРІАЛИ»

В статье рассматривается подход к построению информационной системы управления научными знаниями, представленными в виде оцифрованных электронных документов различных типов. Описан алгоритм автоматической классификации документов по научным направлениям в предметной области «Сверхтвердые материалы».
Ключевые слова: информационная система, автоматическая классификация, электронные документы.

The article discusses the approach to building information management system of scientific knowledge presented in the form of digitized electronic documents of various types. An algorithm for automatic classification of documents on scientific fields in the subject area "Superhard materials.".
Key words: information system, automatic classification, electronic documents.

У статті розглянуто підхід до побудови інформаційної системи керування науковими знаннями, представленими у вигляді оцифрованих електронних документів різних типів. Описано алгоритм автоматичної класифікації документів за науковими напрямами у предметній галузі «Надтверді матеріали»..
Ключові слова: інформаційна система, автоматична класифікація, електронні документи.

В настоящее время в Институте сверхтвердых материалов (ИСМ) создаётся знаниеориентированная информационная система (ЗОИС) [1]. Причиной создания послужил большой объем слабо систематизированных научных данных, представленных в виде электронных копий отчетов по темам, текстов диссертаций, патентов и статей.

Построение данной системы условно можно разделить на три этапа:

- разработка классификатора научных направлений, к которым могут относиться электронные документы.

- разработка программного обеспечения для хранения и автоматической классификации документов (допускающей изменение класса документа в ручном режиме).

- заполнение базы данных электронными документами (в том числе и оцифровка бумажных оригиналов).

В статье освещены первые два этапа, задача оцифрования архива бумажных оригиналов и внесения их в базу не рассматривается.

В институте разработана двухуровневая структура базы данных онтологии научной деятельности ИСМ [2], которая состоит из научных разделов, подразделов и научных направлений. Первый уровень структуры включает направления научно-исследовательской работы ИСМ (утверждены Президиумом НАН Украины). Второй уровень структуры включает основные направления научной деятельности ИСМ.

Разработана трехуровневая структура базы знаний и онтологий предметной области «Сверхтвердые материалы». Подобранные ключевые слова, наиболее корректно характеризуют направления научной деятельности института в целом и научных подразделений.

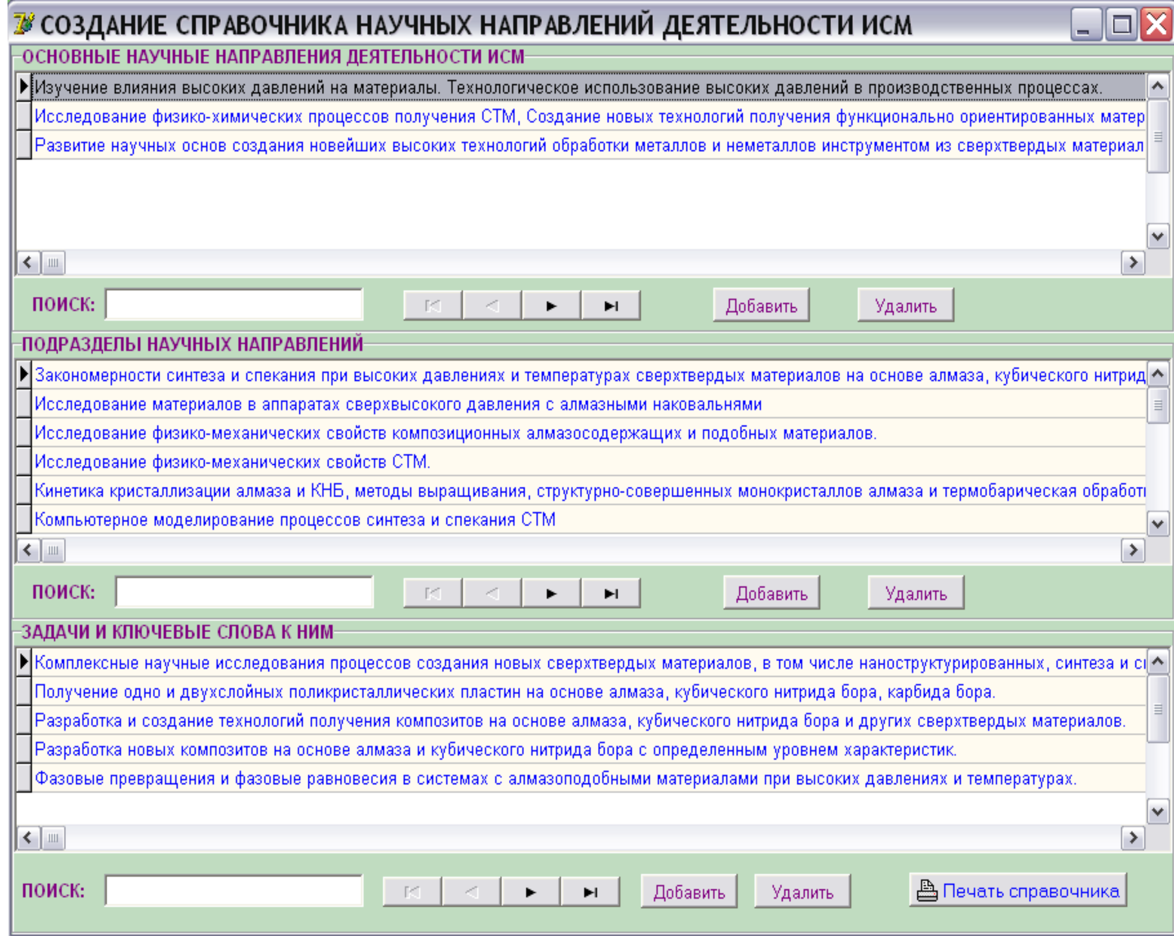


Рис.1. Внешний вид классификатора научных направлений.

Для каждого отдела и лаборатории института был создан перечень основных научных направлений, подразделов научных направлений и научных задач. Данный список был обсужден с сотрудниками отделов и лабораторий, обобщен и переформулирован с целью описания наиболее полно и точно каждой научной задачи.

Таким образом, мы классифицировали все научные работы, проводимые в ИСМ по 3 направлениям:

1. Изучение влияния высоких давлений на материалы. Технологическое использование высоких давлений в производственных процессах.
2. Исследование физико-химических процессов получения СТМ, Создание новых технологий получения функционально ориентированных материалов и изделий из них.
3. Развитие научных основ создания новейших высоких технологий обработки металлов и неметаллов инструментом из сверхтвердых материалов, разработка методов и технологий применения функционально ориентированных материалов.

Научные направления делятся на подразделы (от 5 до 15, в зависимости от направления, суммарно 30), каждый из них, в свою очередь, состоит из научных задач. Подраздел может состоять из 1 или нескольких задач (до 10, суммарно перечислено 111 научных задач, каждая из которых описана характеризующими её ключевыми словами и словосочетаниями). (Рис.1).

Следующим этапом в построении ЗОИС является разработка программного обеспечения для хранения и автоматической классификации документов.

Основой программного обеспечения является СУБД, которая управляет всеми процессами системы и поддерживает гибкий программный интерфейс между базами. Базы ЗОИС содержат структуру главной базы story и вспомогательных реляционных баз avtor, otdel, slova, story_avtor, story_napr. Главная база знаний story ориентирована на размещение информации о научных публикациях ИСМ, научно-технических отчетов, патентов, авторских свидетельств, диссертаций научно-технического архива ИСМ. Программа управления базами знаний предоставляет программный сервис не только для быстрого внесения необходимой информации в главную базу знаний story, но и для создания связки между ней и внешними файлами, содержащими сами тексты научных документов, фотографии, рисунки, схемы, таблицы.

Основной базой данных является база научных документов, которая включает в себя научные отчеты, патенты, диссертации, авторские свидетельства и публикации. Основную роль в данной системе играет вышеописанный справочник научных направлений деятельности ИСМ, который состоит из трех уровней: научных разделов, научных подразделов и научных задач (направлений). Каждое научное направление определяется набором ключевых слов, которые характеризуют это научное направление и, соответственно, относящийся к нему подраздел и раздел. Справочник научных направлений представлен в программной системе в виде трех взаимосвязанных между собой баз данных. В настоящей ЗОИС разработана программная модель классификации электронных документов на основе теории распознавания образов. Образами в системе ЗОИС выступают перечисленные выше виды научных документов, которые описываются названием, авторами, ключевыми словами, рефератами, аннотациями и другими характеристиками. Любое научное направление справочника научных направлений, которое характеризуется общими свойствами для всех его элементов, играет роль класса образов.

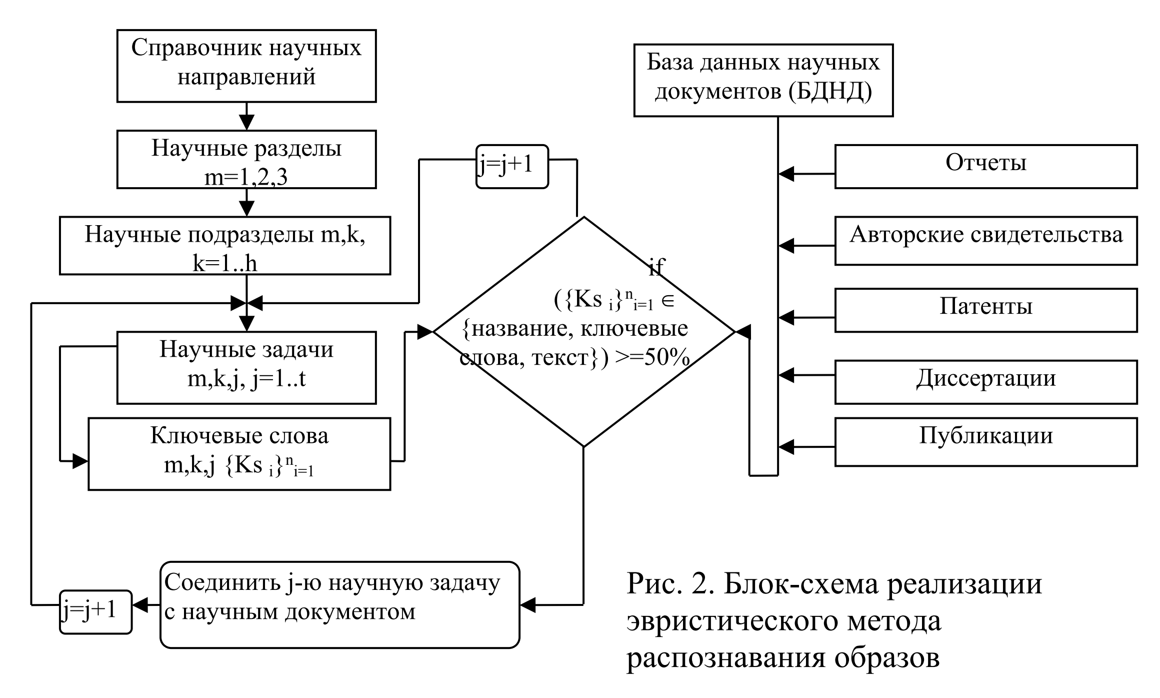


Рис. 2. Блок-схема реализации эвристического метода распознавания образов

В ЗОИС программно реализован эвристический метод распознавания образов, который основан на сравнении неизвестных образов с эталонами. Точнее, если какой-либо входной образ наиболее соответствует эталону i-класса образов, чем какому-либо другому эталону другого класса, то входной образ классифицируется как принадлежащий к i-му классу образов. В системе ЗОИС осуществлена программная реализация сопоставления подобия. Критерием выступает 50% и более вхождение набора ключевых слов (эталона) каждого научного направления (класса) в описание характеристик научных документов (образов). Если условия критерия выполняются, то программная система организует ссылочную связь или привязку данного научного документа к выбранному научному направлению. При этом научный документ может относиться к нескольким научным направлениям. Блок-схема программы, реализующей этот критерий, представлена на рисунке 2.

Практически же программная система для каждой записи в основной базе данных, содержащей все сведения о научном документе, открывает в соответствующей дочерней реляционной базе записи с информацией о том или ином научном направлении, к которому относится данный научный документ (рисунок 3). Так организуется прямая ссылочная связь между документом и научными направлениями из справочника научных направлений, с одной стороны. При этом в дочернюю реляционную базу записывается информация о данном научном документе с другой стороны. Такая программная взаимосвязь между справочником научных направлений и основной базой научных документов посредством промежуточной базы данных реализована в программной системе ЗОИС для удобства осуществления поисковых операций в данной системе. Если в дальнейшем осуществлять поиск по справочнику научных направлений, то программный интерфейс будет давать возможность пользователю сразу находить научные документы, относящиеся к тому или иному научному направлению, и иметь доступ к ним (рисунок 4).

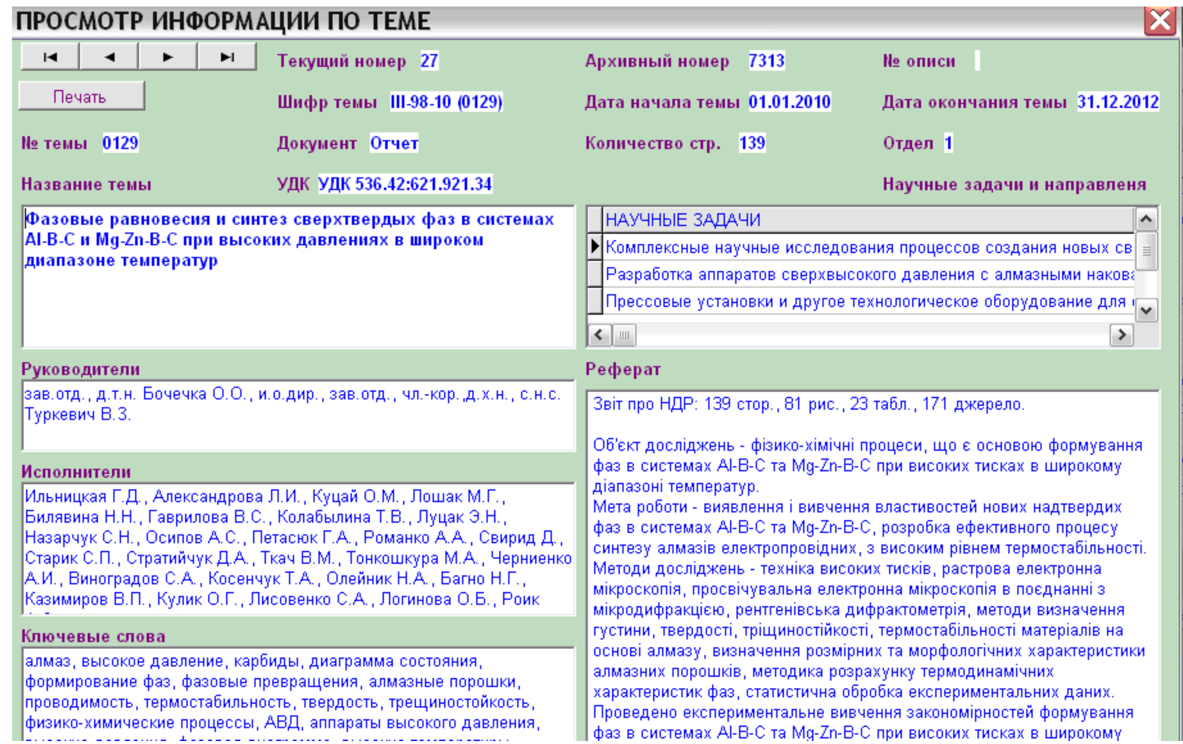


Рис.3. Информация о научном документе в основной базе данных.

Если условия критерия не выполняются, возможно отнесение документа вручную к какому-либо классу. При этом набор ключевых слов задачи пополняется ключевыми словами документа.

Таким образом, в программной системе ЗОИС реализован механизм распределения научной документации по научным направлениям деятельности института.

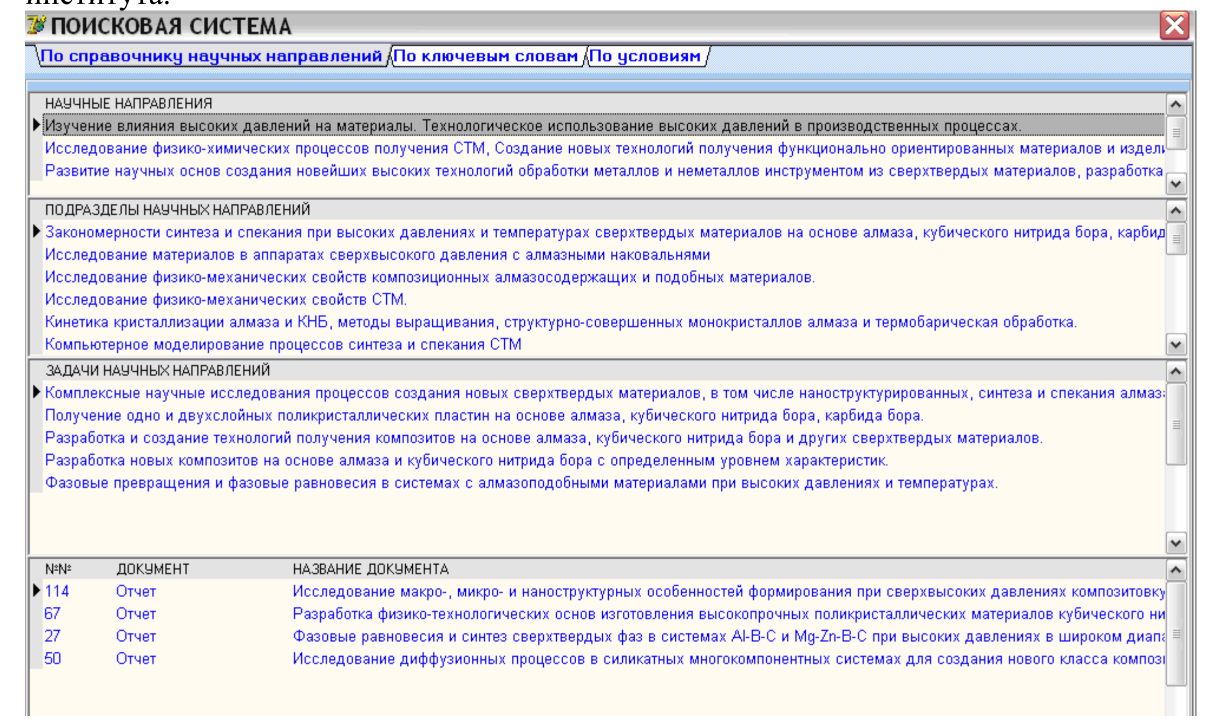


Рис. 4. Поиск документов по справочнику научных направлений.

Литература

1. Кулаковский В.Н., Колодницкий В.Н., Чистяков Е.М., Цегельнюк В.В.Тезаурус как основа построения онтолого-тезаурусной системы управления знаниями в предметной области "Сверхтвердые материалы" / Науково-технічна конференція "Штучний інтелект. Інтелектуальні системи - 2013", 23-27 вересня, п. Кацивелі, Крим: Тезисы докладов. —2013.—Т.0.—С.223-226.
2. Лебедева А. А., Гордашник К. З., Колодницкий В. Н., Чистяков Е. М., Кулаковский В.Н., Сороченко Т. Е., Скворцов И. В., Цегельнюк В. В., Дубенко М. В. Представление и аналитическая обработка данных в предметной области «СТМ» на основе онтолого-тезаурусного анализа // Породоразрушающий и металлообрабатывающий инструмент – техника и технология его изготовления и применения: Сборник научных трудов. – Вып. 17. – Киев: ИСМ им. В. Н. Бакуля НАН Украины, 2014. – С. 370–379.

Literatura

1. Kulakyvskiy V.N. Kolodnitsky V.N. Chistyakov E.M. Tsegelnyuk V.V. Tezaurus as the basis for building ontology-thesaurus knowledge management system in the subject area "superhard materials"/ Scientific and technical conference "Artificial intelligence . Intelligent Systems - 2013 ", September 23-27, p. Katsyveli Crimea: Abstracts. -2013.-T.0.- pp.223-226
2. Lebedeva A.A., Gordashnik K.Z., Kolodnitsky V.N. Chistyakov E.M., Kulakyvskiy V.N, Sorochenko T.E, Skvortsov I.V., Tsegelnyuk V.V., Dubenko M.V. Presentation and analytical data processing in the subject area "SHM" on the basis of ontology, thesaurus analysis // rock cutting and metalworking tools - machinery and technology of its manufacture and use: Collection of scientific works. - Vol. 17. - Kiev: V.N. Bakul`s institute NASU, 2014. - pp. 370-379.

RESUME

V.N. Kulakovskii, V.V. Tsegelnyuk, I.V. Skvortsov.

Management Information System of scientific data in the subject area "superhard materials"

This article describes the stages of building knowledge-oriented information system of the Institute. We consider two stages of development - development of the classifier of electronic documents on scientific directions and the development of software for the storage and automatic document classification.

Classifier of electronic documents is based on a previously created domain ontology "superhard materials". The classifier lists more than a hundred scientific tasks, each of which is described characterizing its keywords and key phrases.

The software is developed using Delphi and MySQL. The algorithm of automatic classification of documents based on a comparison of keywords describing the research area and the words contained in the title and description of the document, and in the full text. The document can be attributed to a number of scientific problems. Manual classification and correction also possible. Also, the algorithm of automatic classification of scientific paper was presented.

В.М. Кулаківський, В.В. Цегельнюк, І.В. Скворцов

Інформаційна система керування науковими даними у предметній галузі «Надтверді матеріали»

У статті описано етапи побудови знанняорієнтованої інформаційної системи інституту. Розглянуто два етапи розробки - розробка класифікатора електронних документів за науковими напрямами та розробка програмного забезпечення для зберігання і автоматичної класифікації документів.

Класифікатор електронних документів розроблений на основі раніше створеної онтології предметної галузі «надтверді матеріали». У класифікаторі пераховано більше сотні наукових завдань, кожне з яких описано за допомогою ключових слів та словосполучень, що максимально чітко це завдання характеризують.

Програмне забезпечення розроблене на Delphi з використанням MySQL. Алгоритм автоматичної класифікації документів базується на зіставленні ключових слів, що характеризують науковий напрям і слів, що містяться як у назві та описі документа, так і повному тексті. Документ може бути віднесений до кількох наукових завдань. Передбачена ручна класифікація і корекція.

В.Н. Кулаковский, В.В. Цегельнюк,И.В. Скворцов

Информационная система управления научными данными в предметной области «Сверхтвердые Материалы»

В статье описаны этапы построения знаниеориентированной информационной системы института. Рассмотрены два этапа разработки – разработка классификатора электронных документов по научным направлениям и разработка программного обеспечения для хранения и автоматической классификации документов.

Классификатор электронных документов разработан на основе ранее созданной онтологии предметной области «сверхтвердые материалы». В классификаторе перечислено более сотни научных задач, каждая из которых описана характеризующими её ключевыми словами и словосочетаниями.

Программное обеспечение разработано на Delphi с использованием MySQL.

Алгоритм автоматической классификации документов базируется на сопоставлении ключевых слов, характеризующих научное направление и слов, содержащихся как в названии и описании документа, так и полном тексте. Документ может быть отнесен к нескольким научным задачам. Предусмотрена ручная классификация и коррекция.

Поступила в редакцию 25.08.2015

